

Primary Structure of the Leukocyte Function-associated Molecule-1 α Subunit: an Integrin with an Embedded Domain Defining a Protein Superfamily

Richard S. Larson, Angel L. Corbi, Lisa Berman, and Timothy Springer

Center for Blood Research, Department of Pathology, Harvard Medical School, Boston, Massachusetts 02115

Abstract. The leukocyte function-associated molecule 1 (LFA-1, CD11a/CD18) is a membrane glycoprotein which functions in cell-cell adhesion by heterophilic interaction with intercellular adhesion molecule 1 (ICAM-1). LFA-1 consists of an α subunit ($M_r = 180,000$) and a β subunit ($M_r = 95,000$). We report the molecular biology and protein sequence of the α subunit. Overlapping cDNAs containing 5,139 nucleotides were isolated using an oligonucleotide specified by tryptic peptide sequence. The mRNA of 5.5 kb is expressed in lymphoid and myeloid cells but not in a bladder carcinoma cell line. The protein has a 1,063-amino acid extracellular domain, a 29-amino acid transmembrane region, and a 53-amino acid cytoplasmic tail. The extracellular domain contains seven repeats. Repeats V-VII are in tandem and contain puta-

tive divalent cation binding sites. LFA-1 has significant homology to the members of the integrin superfamily, having 36% identity with the Mac-1 and p150,95 α subunits and 28% identity with other integrin α subunits. An insertion of ~ 200 amino acids is present in the NH_2 -terminal region of LFA-1. This "inserted/interactive" or I domain is also present in the p150,95 and Mac-1 α subunits but is absent from other integrin α subunits sequenced to date. The I domain has striking homology to three repeats in human von Willebrand factor, two repeats in chicken cartilage matrix protein, and a region of complement factor B. These structural features indicate a bipartite evolution from the integrin family and from an I domain family. These features may also correspond to relevant functional domains.

THE leukocyte function-associated 1 (LFA-1)¹ molecule is a member of a family of three leukocyte glycoproteins involved in cell-cell adhesion. This family of proteins, LFA-1, Mac-1, and p150,95, are heterodimers consisting of distinct α subunits ($M_r = 180,000$, 170,000, and 150,000, respectively) and a common β subunit ($M_r = 95,000$) (25, 47). NH_2 -terminal sequencing of the α subunits in the mouse and human has suggested that they are structurally related (33, 48); however, the cell surface expression and function of these molecules differs. LFA-1 is expressed on virtually all leukocytes and is involved in a large number of adhesion-dependent phenomena. mAb directed against LFA-1 inhibit antigen-specific T helper cell function and cytolytic functions such as cytotoxic T lymphocyte-mediated killing, antibody-dependent cytotoxicity by granulocytes, and natural killer activity (47). LFA-1 is also involved

in antigen-independent interactions that mediate cell localization to sites of inflammation such as leukocyte adhesion to endothelial cells, fibroblasts, epidermal keratinocytes, and synovial cells (12-14, 17, 30, 53). Mac-1 and p150,95 are expressed on monocytes, granulocytes, and some activated lymphocytes, and function as adhesion molecules in cell-cell and cell-substrate interactions as well as complement receptors for C3bi (1).

The importance of these three glycoproteins is signified by the clinical syndrome known as leukocyte adhesion deficiency (LAD) (1). The primary defect in LAD occurs in the β subunit common to LFA-1, Mac-1, and p150,95 (22) resulting in deficient surface expression of the $\alpha\beta$ complexes. LAD patients have recurrent bacterial infections which are sometimes fatal and their leukocytes are deficient in a wide range of adhesion-dependent functions.

The structure of the β subunit common to LFA-1, Mac-1, and p150,95 has been determined (23, 27) and has revealed homology to extracellular matrix (ECM) receptors. These similarities led to the concept of a family of α/β heterodimers designated the integrins (21, 41). The term "integrin" emphasizes the role of these proteins as transmembrane links between the extracellular environment and the cytoskeleton.

1. *Abbreviations used in this paper:* CMP, cartilage matrix protein; ECM, extracellular matrix; FNR, fibronectin receptor; ICAM-1, intracellular adhesion molecule-1; LFA, leukocyte function-associated 1; vWF, von Willebrand factor.

Three subfamilies of integrins are defined by their distinct β subunits. The β_1 subunit is common to the fibronectin receptor (FNR) and some antigens appearing very late in leukocyte activation whereas β_3 is common to platelet glycoprotein IIb/IIIa (gpIIb/IIIa) and the vitronectin receptor. β_1 and β_3 integrins are all ECM receptors and are involved in cell-substrate adhesion, matrix assembly, regulation of cell growth, differentiation, and localization during morphogenesis, and wound healing. We will refer to the β_1 and β_3 family as the extracellular matrix receptor integrins. The β_2 subunit is common to LFA-1, Mac-1, and p150,95, whose expression is limited to leukocytes, and we designate these as the leukocyte integrins.

The ligand of LFA-1 is an inducible cell surface glycoprotein, intercellular adhesion molecule-1 (ICAM-1) ($M_r = 90,000$), which is found on cells of many lineages including leukocytes, endothelial cells, fibroblasts, and epidermal keratinocytes (14, 29, 40, 46, 49). ICAM-1 mRNA and surface expression is induced by inflammatory mediators including interferon- γ , interleukin-1, tumor necrosis factor, and lipopolysaccharide (36, 46, 49); thus, it may regulate cell interaction and localization in inflammation. LFA-1-dependent adhesion of cells to planar lipid membranes containing ICAM-1 requires metabolic energy, a functional cytoskeleton, and divalent cations (29). Cell activation can enhance LFA-1-dependent adhesion without any effect on LFA-1 or ICAM-1 surface expression as shown by phorbol ester-induced homotypic adhesion of B, T, and monocytic cells (39). ICAM-1 is a member of the immunoglobulin superfamily and consists of five immunoglobulin constant region-like domains. In contrast to most ligands of the ECM receptor integrins which contain the critical recognition sequence arginine-glycine-aspartic acid (RGD), ICAM-1 does not contain an RGD sequence (46, 49). The LFA-1-ICAM-1 receptor-ligand pair is thus far the only known example of a member of the integrin superfamily interacting with a member of the immunoglobulin superfamily.

We have been interested in the structural basis for the important function of LFA-1 in inflammation and the immune response. Furthermore, we wished to define the relationship of LFA-1 to other leukocyte integrins and to the ECM receptor integrins. The lack of an RGD sequence in ICAM-1 raised the question of whether the LFA-1 α subunit has structural features typical of ECM integrins or whether novel features consistent with the lack of RGD recognition exist. Therefore, we have characterized the LFA-1 α subunit at the protein and mRNA level and have determined its complete sequence. The amino acid sequence demonstrates that LFA-1 is an integral membrane protein containing an extracellular domain, a single hydrophobic transmembrane domain, and a cytoplasmic tail. The extracellular domain has seven repeats; the three repeats located most COOH-terminal contain putative divalent cation binding sites. LFA-1 has significant homology with the other members of the integrin superfamily. A region near the NH_2 terminus of the molecule contains an insertion of ~ 200 amino acids similar to p150,95 (9) and Mac-1 (4, 7, 38). This domain has significant homology to the type A domains of von Willebrand factor (vWF), complement factor B, and the repeats of chicken cartilage matrix protein (CMP). These similarities suggest relevant functional domains within the LFA-1 α subunit as well as novel evolutionary relationships.

Materials and Methods

Protein Purification

The mAb, TSI/22, which is directed against the LFA-1 α subunit, was purified and coupled using cyanogen bromide to CL-4B Sepharose at 2 mg mAb per ml of packed bed. SKW3 cells (42.2 g) were lysed in 300 ml of lysis buffer (25) and the lysate was spun at 5,000 g, the pellet discarded, and then spun at 16,000 g for 2 h. The supernatant was then sequentially passed through a precolumn of activated and quenched CL-4B Sepharose and then a TSI/22 mAb Sepharose column. The TSI/22 column was washed sequentially (25), and the LFA-1 molecule was eluted with 0.5 M NaCl, 0.1% Triton X-100, 1 mM iodoacetamide, 10 U/ml aprotinin, and 0.025% NaN_3 , 50 mM triethylamine, pH 11.5, and the pH immediately neutralized. The fractions containing LFA-1 were pooled, lyophilized, and precipitated in 5 vol ethanol at $-20^\circ C$ overnight.

Purified protein was reduced and alkylated (23) and subjected to preparative SDS-PAGE. The band corresponding to the α subunit was visualized with 1 M KCl, excised, and electroeluted (20). The purified α subunit was lyophilized and precipitated in 4 vol ethanol at $-20^\circ C$ overnight. The pellet was resuspended and digested with 1% (wt/wt) trypsin (23). The tryptic fragments were then isolated by HPLC (Beckman Instruments Inc., Palo Alto, CA) on a C4 reverse phase column (Vydac, Hesperia, CA). The peptides were eluted on a 0–60% acetonitrile gradient in 1% trifluoroacetic acid. Several peaks were rechromatographed isocratically in a concentration of acetonitrile determined by the equation, $F = 0.9E^{-2}$, where F is the volume percentage of acetonitrile under isocratic conditions for a peptide that eluted at E percent during the linear gradient (57). Peaks were collected in 1.5-ml polypropylene tubes and concentrated to $< 50 \mu l$. Eight peaks were subjected to microsequencing. The sequence of one peptide, L64, was used to synthesize a single sequence oligonucleotide (5'-GGGATGTTGGT-CATGGATGGTGGGCTCAAT-3') according to suggestions of Lathe (26).

cDNA Cloning, Restriction Mapping, and Nucleotide Sequencing

The production and screening of the cDNA library was performed as previously described (9). Restriction maps of the selected clones were determined by double and partial digests (28). Restriction fragments were subcloned into either M13mp18 and mp19 (31) or pGEM-3Z, 4Z, or 7Z (Promega Biotec, Madison, WI). Deletions of the fragments in pGEM were made using Exonuclease III and SI nuclease (18). Sequencing was by the dideoxy termination method (43). The coding, 5'-untranslated and 3'-untranslated regions were determined 100, 100, and 36.7% in both orientations, respectively.

Southern and Northern Blot

Southern blots were performed as described elsewhere (8). For Northern blots, 10 μg of poly (A+) RNA isolated from SKW3, U937, IB4, or EJ cells was subjected to electrophoresis on a 1.0% formaldehyde gel and transferred to nitrocellulose (5). The nitrocellulose (Bio-Rad Laboratories, Richmond, CA) was prehybridized and hybridized in $2\times$ SSC, $1\times$ Denhardt's solution, 0.1% SDS, and 10 $\mu g/ml$ of herring sperm DNA. A 1.8-kb Eco RI probe from the 5' end of the cDNA clone, λ 3R1 (Fig. 2), was labeled by nick translation and used as probe.

Computer Analysis

Homology searches and alignment of sequences used the Microgenie DNA program (Beckman Instruments Inc.), FASTP (56) on the NBRF and NEW databases (National Biomedical Research Foundation, Washington, DC), and FASTP using the SWISS-PROT data base (Bionet Intelligenetics, Mountainview, CA). These alignments were then optimized, and the percent identity and statistical significance were determined using ALIGN (NBRF) (11).

Hydrophobicity was determined according to Hopp and Woods using the Microgenie DNA program (19).

Results and Discussion

Protein Purification and Peptide Sequence

LFA-1 was solubilized from SKW3, a T lymphoma cell line,

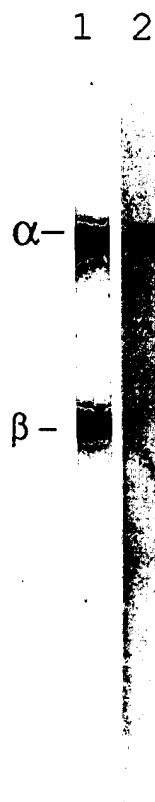


Figure 1. Protein purification of LFA-1 α subunit. Silver-stained SDS-PAGE gel showing a column fraction of the affinity-purified protein (lane 1) and the electroeluted α subunit (lane 2).

and isolated by mAb affinity chromatography using an antibody directed against the α subunit. SDS-PAGE showed the α and β subunits (Fig. 1, lane 1). The α subunit was further purified by preparative SDS-PAGE (Fig. 1, lane 2) and the purified α subunit was digested with trypsin and the peptides were isolated by reverse phase HPLC. The sequence of nine peptides was determined by microsequencing (Table I) and the sequence possessing the lowest codon redundancy was used to specify a single oligonucleotide sequence which used the most commonly occurring human codons (26).

cDNA Cloning and Characterization

The 32-mer oligonucleotide probe was used to isolate 20 clones from a size-selected λ gt10 cDNA library constructed from PMA-stimulated myeloid cells (9). These cells have been previously shown to synthesize the LFA-1 α subunit

Table I. Sequences of Tryptic Peptides

Residues	Amino acid sequence
95-107	X X D Q (N) T Y L S G L (E) Y L F
199-210	H M L L L T N T F G A I
254-260	Y I I G I G K
405-413	V L L F Q E X Q G
494-503	G E A I T A L T X I
541-554	I E G T Q V L S G I Q X F G
564-576	X (L) E (G) D (V/G) L A D V A V G A E
803-817	K V E M L K P H S E I X V S
929-946	I (E/Q) P S I H D H N I P X L E A V X G

Parentheses indicate ambiguity in the sequence. The underlined residues were used to generate an oligonucleotide probe.

(32). The insert size was determined and the longest clone, λ 5L5, was restriction mapped, and sequenced (Fig. 2). This clone contained the nucleotide sequence corresponding to the oligonucleotide probe (85% identity to the "best-guess" probe) and agreed perfectly with the tryptic peptide sequence (16 of 16 residues). However, this clone did not encode the entire protein since not all of the tryptic peptide sequences were present in the open reading frame. The 5' 1.0-kb Eco RI fragment of λ 5L5 was used to select 14 additional clones. The clone λ 3R1 had an identical restriction map in the overlapping regions and contained an additional 1.0-kb 5' fragment (Fig. 2).

The composite sequence of λ 5L5 and λ 3R1 contains 5,139 nucleotides (Fig. 3). There is an open reading frame of 3,510 nucleotides, a 5' untranslated region of 94 nucleotides, and a 3' untranslated region of 1,535 nucleotides which contains a polyadenylation signal site 15 nucleotides before the poly(A⁺) tail. Within the 3' untranslated region there is a typical Alu repeat consisting of two tandem related sequences each terminated by an A-rich segment (16).

Northern blots demonstrate an LFA-1 α subunit mRNA of 5.5 kb in SKW3 T lymphoma cells, U937 myelomonocytic cells, and IB4 B lymphoblastoid cells (Fig. 4 A, lanes 1-3); however, no signal was detected in EJ bladder carcinoma cells (Fig. 4 A, lane 4). The mRNA expression is in agreement with the restriction of cell surface expression of LFA-1 to hematopoietic cells. The β subunit showed the same pattern of expression (Fig. 4 B). In Southern blots, the 1.2-kb 5' Eco RI-Bam HI fragment from clone λ 2L2 hybridized to two fragments of 10 and 8 kb (8). A genomic clone isolated from a cosmid library possesses the same two fragments

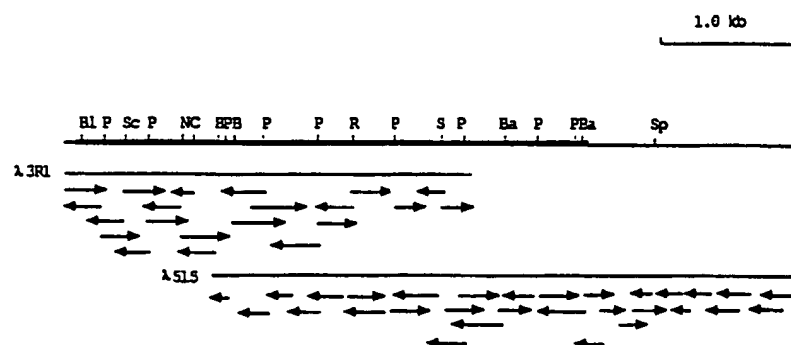


Figure 2. Restriction map of the LFA-1 α subunit cDNA clones. The open reading frame and untranslated regions are indicated with a thick and thin line, respectively. Arrows indicating the sequencing strategy are shown. The relevant restriction sites are Bal I (Bl), Bam HI (Ba), Bgl II (B), Cla I (C), Eco RI (R), Nru I (N), Pst I (P), Sca I (Sc), Sma (S), and Sph I (Sp).

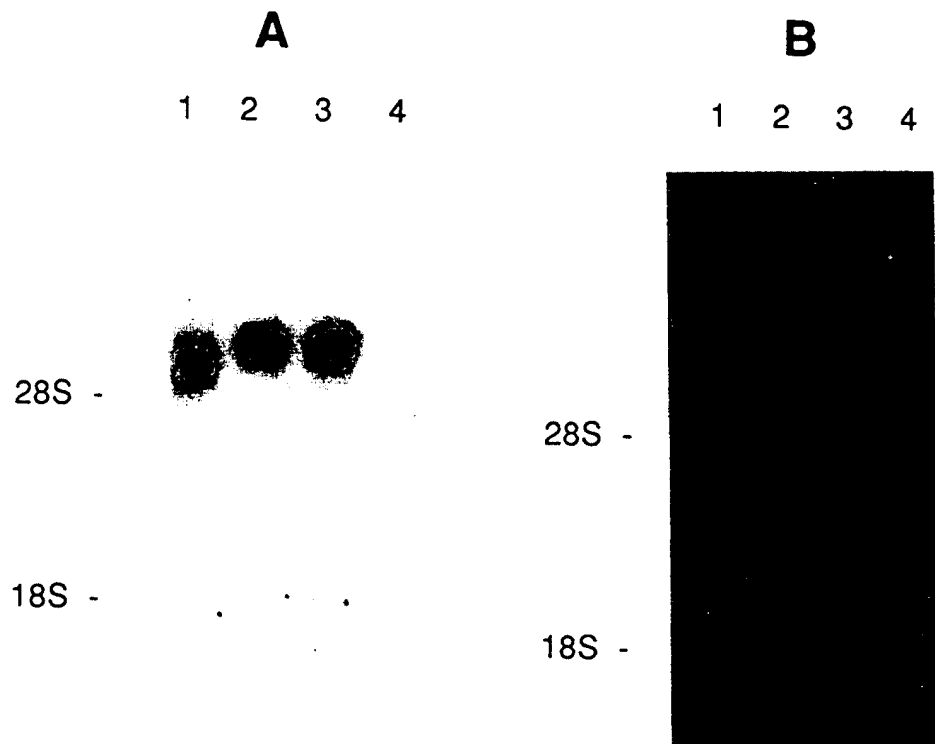


Figure 4. (A) Northern blot analysis of 10 μ g of poly (A⁺) selected RNA from SKW3 (lane 1), U937 (lane 2), IB4 (lane 3), and EJ cells (lane 4). (B) Identical blot reprobed with the β subunit cDNA probe (23). The position of the 28 and 18 S rRNA is indicated.

which are contiguous and hybridize with different regions of the cDNA (unpublished data) showing that the LFA-1 α subunit is a single copy gene.

Protein Sequence

All 105 amino acids determined by microsequencing of tryptic peptides were found in the translated open reading frame, confirming the authenticity of the cDNA clones (Fig. 3). Hydrophobicity analysis shows that the LFA-1 α subunit is a typical transmembrane protein with a 25-residue hydrophobic signal sequence, an extracellular domain of 1,063 residues, a single hydrophobic transmembrane region of 29 residues, and a short cytoplasmic tail of 53 residues. The NH₂-terminal residue of human LFA-1 was identified by homology to the NH₂-terminal sequence of murine LFA-1 (48) (Fig. 6, see below). Human LFA-1 has 55% identity with murine LFA-1 over the first 20 amino acids. A classical signal peptide with a consensus sequence (Ala-X-Ser/Pro) for the cleavage peptidase precedes the NH₂-terminal sequence. There are three putative upstream transcription initiation sites (ATG) in frame. The use of the first initiation site

(nucleotide position 89) is generally favored (24) and gives a 25-residue signal sequence with several NH₂-terminal polar groups as is typically found in signal sequences (54).

The mature protein is $M_r = 126,193$. 12 N-linked glycosylation sites (Asn-X-Thr/Ser) are present in the extracellular domain. These findings are consistent with the size previously determined by SDS-PAGE for the in vitro translated murine LFA-1 α subunit ($M_r = 140,000$), the murine and human LFA-1 α subunit glycoproteins ($M_r = 180,000$ and 177,000) (44), and with previous studies on the glycosylation of LFA-1 (10, 32).

Within the extracellular domain there are seven internal repeats (Fig. 5). The degree of identity is highest among the three repeats (14.5–33%) located toward the COOH terminus which show a statistically significant relationship ($P < 10^{-2} - < 10^{-6}$). The relationship among the four repeats located toward the NH₂ terminus is weaker and is discernible by conservation of flanking sequences (Fig. 5); the homology between repeat IV and V is significant ($P < 10^{-4}$). The central regions of the three COOH-terminal repeats is similar with the EF hand divalent cation binding site motif, perhaps due to convergent evolution (Fig. 5). Previous studies have shown that Mg²⁺ alone or at lower concentration in conjunction with Ca²⁺ is necessary for ligand binding func-

2. These data sequence have been submitted to the EMBL/GenBank Data Libraries under the accession number Y00796.

Figure 3. Nucleotide sequence and derived protein sequence of the α subunit of LFA-1.² The potential glycosylation sites are boxed. The sequences of the tryptic peptides and the transmembrane region are underlined with a thick and shaded line, respectively. The putative serine phosphorylation sites are circled. The nucleotides in the 3' untranslated region that are underlined correspond to an AluI sequence and the polyadenylation site is boxed.

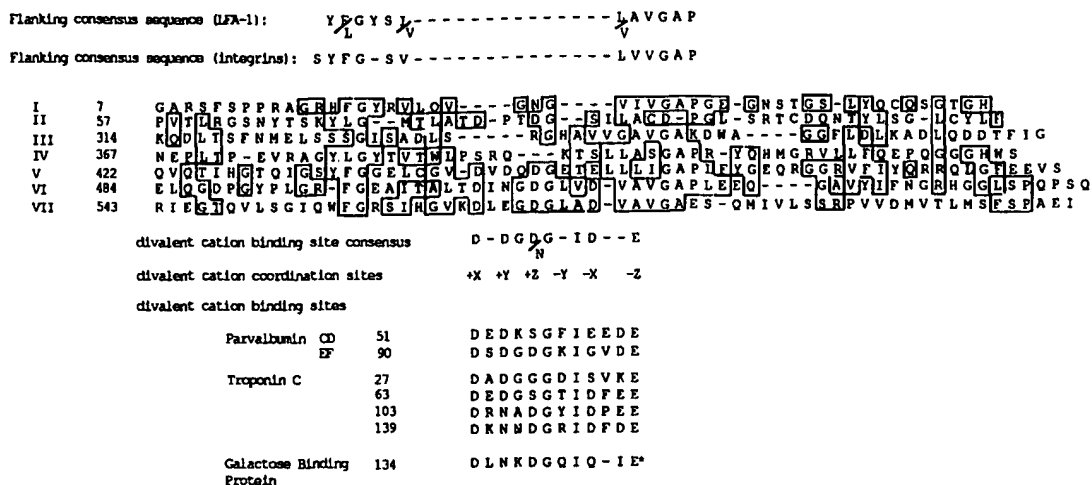


Figure 5. Alignment of the internal repeats of the LFA-1 α subunit. The consensus flanking sequence is shown above the aligned repeats. Consensus residues were defined as those appearing in at least 30% of the integrin α subunits analyzed (Mac-1 [4, 7], p150,95 [9], vitronectin receptor [50], fibronectin receptor [3], and platelet glycoprotein IIb [37]). The calcium and magnesium binding sites of parvalbumin, troponin C (52), and galactose binding protein (55) are aligned as shown. The asterisk indicates that more than one oxygen is involved in cation binding.

tion (29) (Dustin, M. L., and T. Springer unpublished results), suggesting that these sites bind divalent cations.

Distinct Subfamilies of Integrin α Subunits

We compared the LFA-1 α subunit to other integrin α subunits (Fig. 6). The LFA-1 α subunit has striking and consistently higher homology to other leukocyte integrin α subunits (Mac-1, 35.7%; p150,95, 37.4%) than to ECM receptor integrin α subunits (vitronectin receptor, 26.4%; FNR, 27.8%; gpIIb 30.2%). The ECM receptor integrin α subunits are more related to one another (\bar{x} = 41.9% SD = 3.7%) than to the leukocyte integrins (Fig. 8, see below). Further structural features distinguish the leukocyte and ECM receptor integrins. The leukocyte integrins contain an insertion of \sim 200 amino acids near the NH₂-terminal region of the protein that is not present in the three sequenced ECM receptor integrins (Fig. 6). Furthermore, a region containing dibasic amino acid protease cleavage sites in vitronectin receptor, FNR, and gpIIb/IIIa (residues 853–871) is absent from LFA-1 as well as p150,95 and Mac-1 α subunits, correlating with the lack of proteolytic processing of the leukocyte integrin α subunits (42). As a whole, these structural features define two subfamilies of the α subunit integrins, the leukocyte integrins and the ECM receptor integrins.

All the integrin α subunits thus far sequenced have tandem repeats similar to those of LFA-1 (3, 9, 37, 41, 51) (Fig. 5). Putative divalent cation binding sites are present in repeats V–VII in all the integrin α subunits and in repeat IV only in the ECM receptor integrins. These sites are similar to the EF hand divalent cation binding sites located in the turn between two α helices in troponin C, parvalbumin (52), and galactose binding protein (55) among others. In the EF hand loop the metal is chelated to six oxygen-containing side groups spaced every second or third residue. The putative divalent cation binding sites on LFA-1 and the other integrins have a similar primary structure with the putative chelating residues and intervening glycine residues being conserved (Fig. 5). However, the residue in the –Z position is replaced by a hydrophobic residue.

An Inserted Domain in the Leukocyte Integrins

Searches of the NBRF and SWISS-PROT protein sequence data banks revealed that the LFA-1 domain of \sim 200 amino acids, which is not present in the sequenced ECM integrins, is homologous to the domains of the same size in vWF and a cartilage matrix protein (Figs. 7 and 8). These alignments have 20.4–32.1% identity and are statistically significant (P < 10^{-9} – 10^{-23}). Of the integrins sequenced to date, this 200–amino acid domain is unique to the leukocyte integrins. Homologous domains are inserted in several proteins and in well-studied examples have been documented to mediate interaction with ligand (see below). For this reason we will refer to the 200–residue region of the LFA-1, Mac-1, and p150,95 α subunits as the “I” (inserted/interactive) domain. The homology unit is present in three tandem repeats in vWF (A domains) (45) (P < 10^{-15} – 10^{-23} for comparison of the three repeats to one another) and two repeats separated by an EGF domain in CMP (2) (P < 10^{-23}). With the exception of the NH₂-terminal region of the CMP repeat 1 which has not been sequenced, the repeating units in both vWF and CMP correspond precisely to the region homologous to the I domain, supporting the concept that this homology unit is a domain. Similar structural homologies have been noted for the murine and human Mac-1 α subunit (7, 38) and the degree of homology of Mac-1 and p150,95 with vWF and CMP is similar to that found with LFA-1 (17.8–31.6% identity, P < 10^{-9} – 10^{-23}). Factor B has previously been found to be homologous with the vWF A repeats (P < 10^{-4} – 10^{-7}) (45); factor B in turn is homologous but at a lower level with the LFA-1, Mac-1, and p150,95 I domains (P < 10^{-2} – 10^{-4}) (7, 38). In factor B a single homology unit is bounded on one side by the site for the cleavage which activates the B zymogen to the active Bb fragment, and on the other side by the serine protease domain (6).

Based on these homologies, we propose the following pathway for leukocyte integrin evolution. A primordial gene duplicated and gave rise to at least two branches of integrin α subunits, the leukocyte integrins and the ECM integrin receptors. The percent identity among the I domains is very

LFA1	130	VDLVFLDGGSSILQDFQKILDFHRDVF--KRLNTSMQFANVQFSTSYKTEFQFSDYV	
MAC1	133	SDAFPLDGGSSILPHDFRMKEFVSTVMEQLKESAT--QSLMOTSEFRINHTFEFQ	
P150	131	QDVFLLDGGSSITSSRNFTMMNFVRAVTSQFQFST--QSLMOTSNKFOHTFEFQ	
VWFA1	513	LQVFLDGGSSRLSEAEFEVLKAFVVDKFERLRI--SQKWVR--NAVVEYHDSHATIGLK	
VWFA2	734	LQVAVLDGSSDKIGEAADNRSKEFHEEDTORMDV--GQDSIH--VTVLQYSYMTVEYFPS	
VWFA3	927	LQVILLDGGSSFPASYFDEMKSFAKAYISKANIGPRLTQ--VSVLQYGSITTIIDVPWN	
FACTOR B	244	MNIYLVLDGSSSTGASNTGAKKCLVNLIEKVASYGVKPRYGLVTYADYFPIWVKVSEAD	
CMP2	195	LQVFLDGGSSVRRFENTELVKKKFINQ--IVESLEV--SEKQNDVGLVQYSSSV--RQEFPL	
CMP1	1	-----MGVINYASAV--KNEFSL	
LFA1		KALQFDALLRHKHILL-----TNTFGAI--NVVATEVFREELGARPDATKVI-----	
MAC1		NNPMERSLQPTQILGR-----TDTATGIT--NVVRELFINITNGARKNAFMT-----	
P150		RTSNELSLASVHQLQGF-----TDTATAIQ--NVVHRLPHASYGARPDATKVI-----	
VWFA1		DRKPESELRRIASQVKYAG-----SQVQSTSEVLKYILPFIKIDREASRIA-----LL	
VWFA2		EAQSKGDIQRRVREIRYQGGNR--TNTGLAL--RVLSDHSEVLVSQDRECAPNLV-----Y	
VWFA3		VVPEKAHLCLSLVDVQREGGPSQIGDALGFVAV--RVLTSF-----GARPCASRAV-----VI	
FACTOR B		SS--NAQWVTRQLNEINYEHLKLSGTNTKKALQAVYSMMSPDDVPPEGWNRTRH-----VII	
CMP2		GQFKNKDKIAA--VKKMAVMEG-----TNTGQAL--KVLVDSSPSIANGARPGVPRVG-----II	
CMP1		RTHTQTKAE--LLQAVARIEPLSTG--TNTGLAI--QFAISRAISDTEGARLRSFNINKVAI--	
LFA1		ITDGE--ATDS--GNIDA--AKD--TIRYHIGICKHFOIKESQET--CHKFASKPASEFVK	
MAC1		VITDGEKFGDPLGYEDVPEADREGVIRYVIGVGDAPFSEKSRQE--LNTIASKPRDHVF	
P150		VITDGEKKEGDSLQYKQVIPMAAAGTIRYVIGVGLAFQNRMSWKE--LNDIASKPSCCHIF	
VWFA1		LMASQEPQRMERNFVRVYVQGLKKKKVIVIPVIGIPHANIRQIR--LIEKQA--PENKAFV	
VWFA2		MVIGENPASDEKRL--PGDIQ-----VVPTIGVGNANVCE-----LERIGW--FNAP--	
VWFA3		LVLIV--SVDSVDAADANARS-----NRVTVPFPIGIGORYDAAG-----LAILAG--PAGDSNV	
FACTOR B		LMTDGLNHNMGDFPITVIDE-----TROLLYIGKORKNPR--EDYLDVYVFGVGP--LGN	
CMP2		VITDGRSQDYIT--DAKKAKDLGFRHFAVGIVNAV-----EDELREIASEPVVHYF	
CMP1		VITDGRPDQGVQ--VSARARQAGTEIF--ALGVGRVD-----MHTLROIASEPLDDHVD	
LFA1		ITDT--FEKCKDLFTELOKIMVIEGTSKODLTSFMDESSS	327
MAC1		QVNN--FEALKTIQNLREKIFAIEGTOTGSSSSFEHMSQE	337
P150		KVED--FDALKTIQNLREKIFAIEGTETTSSSFELEMAQE	335
VWFA1		LSS--VDE--LEQQRDEIVSYLCDLAPEAPPPTLPFDMAQVTVGP	717
VWFA2		ITLQDFETLPREAPDLVLRQ--CSGEGILIPPLSPAPD	922
VWFA3		VKLQRIEDLPMTVNLGN-----SFLHKCSGFVRI	1115
FACTOR B		QVNINALASK--KDNEQHVFK--VKMDENLEDFYQMIDESQL	450
CMP2		YTA--DRTISNIGKKLQKLCVSEDFCECKSVKFKTKV	393
CMP1		VVESYSVIELDTHKGFDEAFQVSDLCATGDHDEQEQICI	172

Figure 7. Alignment and comparison of the LFA-1 α subunit I domain with the homologous domains in Mac-1 (4, 7), p150,95 (9), vWF (45), factor B (34), and CMP (3).

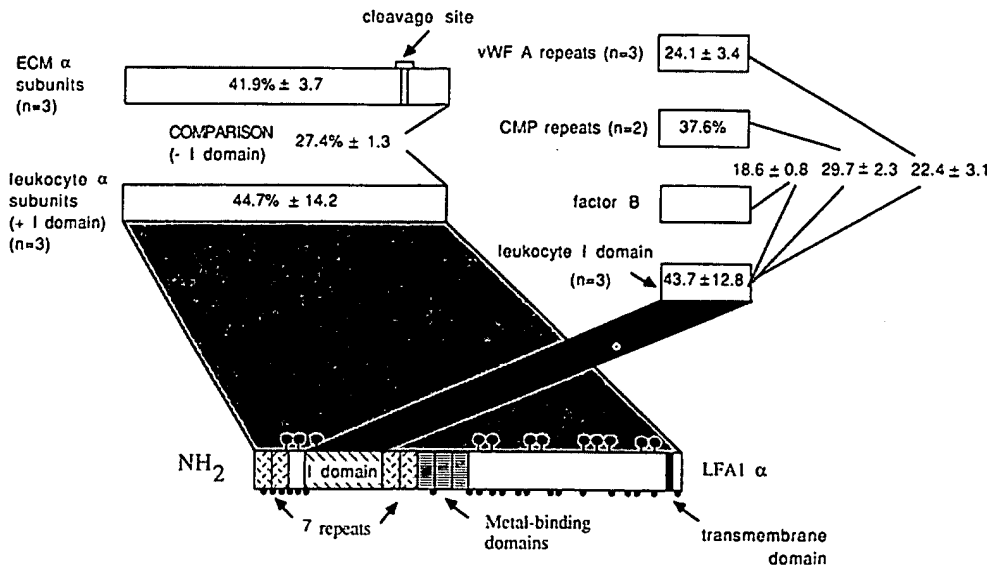
similar to the overall percent identity among the leukocyte integrins, implying that a single I domain incorporated into a primordial leukocyte integrin only once rather than independently. Then, the gene duplicated and gave rise to LFA-1 and a Mac-1/p150,95 primordial gene. Further duplication of the Mac-1/p150,95 primordial gene gave rise to Mac-1 and p150,95. This scheme is consistent with the observation that Mac-1 and p150,95 are more closely related to each other (61.0%) than to LFA-1 (35.7, 37.4%, respectively). Since LFA-1, Mac-1, and p150,95 α subunits are located on chromosome 16 band p11 (8), these α subunits genes remained in close proximity. ECM α subunits appear to be a separate branch since they are more closely related to each other than to the leukocyte integrins even ignoring the I domain (Fig. 8). Although three sequenced α subunits of ECM integrins bind to two different β subunits, the two α subunits associated with the β_3 subunit are not more closely related to one another than to the α subunit associated with the β_1 subunit. Thus, although there are three distinct β subunits, only two subfamilies of α subunits can be distinguished by sequence homology considerations. The structural properties of the leukocyte integrin α subunits may not be unique to their subfamily since some VLA integrin α subunits may have a homologous I domain and lack a proteolytic cleavage site (Hemler, M. personal communication).

The I domain appears to be incorporated in proteins which are otherwise structurally distinct. These structurally simi-

lar domains form an important protein superfamily which we will designate the I domain superfamily. Repeats within vWF and within CMP are internally more related than to the domains in other proteins (Fig. 8). Thus, it appears a single domain was incorporated and then duplicated or triplicated. Members of the I domain superfamily serve important recognition functions in several proteins. The A1 domain of vWF binds to glycoprotein Ib and heparin (15) while both the A1 and A3 domains are involved in binding to collagen. The domain in factor B is located in a region of the molecule available for interaction with its ligand C3b (6, 34). The domain in CMP may also have an important role in interaction with collagen (2) and cartilage proteoglycan (35). These domains lack glycosylation and cysteines for the most part. Similarly, the I domain and the repeats with divalent cation binding sites in LFA-1 contain only one N-glycosylation site and cysteine (Fig. 8). Similar to the Mac-1 (7, 38) and p150,95 (9) α subunits, the N-glycosylation sites and cysteines are located primarily NH₂-terminal and COOH-terminal to this region. Therefore, the I domain would be accessible to ligand and capable of conformational changes which may be important in regulation of ligand binding (39).

Structural differences between the ECM receptor integrin α subunits and LFA-1 correlate with differences in recognition specificity; RGD containing peptides block the binding of the VNR, FNR, and gpIIb/IIIa to their ligands (41) but not binding of LFA-1 to ICAM-1 (29). Furthermore, ICAM-1

Figure 8. Schematic representation of the evolutionary relationships of the domains homologous to the I domain. Black lollipops (above schematic) and circles (below schematic) indicate the respective sites of potential glycosylation and cysteines within LFA-1. The percent identity and the standard deviations among the different integrin α subunits and members of the I domain superfamily are shown. Sequences were compared pairwise using the ALIGN program. The I domain sequences were as follows: LFA-1, Mac-1, and p150,95 α subunits (130–327; 133–337; 131–335, respectively), the A repeats of von Willebrand factor (A1: 513–717; A2: 734–922; A3: 927–1115), factor B (245–450), and collagen matrix protein (repeat 1: 1–167; repeat 2: 195–393). Relevant probability values are shown in the text.



does not contain an RGD sequence and unlike other known integrin ligands, is a member of the immunoglobulin gene superfamily. It will be of interest to determine whether the I domain confers specificity for non-RGD containing ligands.

This study demonstrates that the LFA-1 α subunit belongs to the integrin superfamily but possesses an additional domain. This I domain and homologous domains constitute a protein "domain" family that is of functional importance. Since LFA-1 is involved in a large number of leukocyte functions and may have more than one ligand (39) (Dustin, M. L., and T. Springer, manuscript in preparation), it is possible the more than one functional domain exists in the LFA-1 α subunit. The availability of cDNA clones for the α and β subunits will allow these and other structure-function relationships to be examined.

We thank Dr. D. Giard and the Massachusetts Institute of Technology cell culture center for growth of SKW3 cells. We thank Perisamy Selvaraj for his assistance with the LFA-1 purification, and Don Faulkner for his computer graphics assistance.

This work was supported by National Institutes of Health grant CA 31798.

Received for publication 1 August 1988, and in revised form 22 September 1988.

References

- Anderson, D. C., and T. A. Springer. 1987. Leukocyte adhesion deficiency: an inherited defect in the Mac-1, LFA-1, and p150,95 glycoproteins. *Annu. Rev. Med.* 38:175–194.
- Argaves, W. S., F. Deak, K. J. Sparks, I. Kiss, and P. F. Goetinck. 1987. Structural features of cartilage matrix protein deduced from cDNA. *Proc. Natl. Acad. Sci. USA* 84:464–468.
- Argaves, W. S., S. Suzuki, H. Arai, K. Thompson, M. D. Pierschbacher, and E. Ruoslahti. 1987. Amino acid sequence of the human fibronectin receptor. *J. Cell Biol.* 105:1183–1190.
- Arnaout, M. A., S. K. Gupta, M. W. Pierce, and D. G. Tenen. 1988. Amino acid sequence of the alpha subunit of human leukocyte adhesion receptor Mol (complement receptor type 3). *J. Cell Biol.* 106:2153–2158.
- Ausubel, F. M., R. Brent, R. Kingston, D. D. Moore, J. A. Smith, J. G. Seidman, and K. Struhl. 1987. *Current Protocols in Molecular Biology*. Green Publishing Associates and Wiley-Interscience, NY. 12.8.9 pp.
- Bentley, D. R. 1986. Primary structure of human complement component 2. *Biochem. J.* 239:339–345.
- Corbi, A. L., T. K. Kishimoto, L. J. Miller, and T. A. Springer. 1988. The human leukocyte adhesion glycoprotein Mac-1 (Complement receptor type 3, CD11b) alpha subunit: cloning, primary structure, and relation to the integrins, von Willebrand factor and factor B. *J. Biol. Chem.* 263:12403–12411.
- Corbi, A. L., R. S. Larson, T. K. Kishimoto, T. A. Springer, and C. C. Morton. 1988. Chromosomal location of the genes encoding the leukocyte adhesion receptors LFA-1, Mac-1 and p150,95. Identification of a gene cluster involved in cell adhesion. *J. Exp. Med.* 167:1597–1607.
- Corbi, A. L., L. J. Miller, K. O'Connor, R. S. Larson, and T. A. Springer. 1987. cDNA cloning and complete primary structure of the alpha subunit of a leukocyte adhesion glycoprotein. p150,95. *EMBO (Eur. Mol. Biol. Organ.) J.* 6:4023–4028.
- Dahms, N. M., and G. W. Hart. 1985. Lymphocyte function-associated antigen 1 (LFA-1) contains sulfated N-linked oligosaccharides. *J. Immunol.* 134:3978–3986.
- Dayhoff, M. O., W. C. Barker, and L. T. Hunt. 1983. Establishing homologies in protein sequences. *Methods Enzymol.* 91:524–545.
- Dustin, M. L., and T. A. Springer. 1988. Lymphocyte function-associated antigen-1 (LFA-1) interaction with intercellular adhesion molecule-1 (ICAM-1) is one of at least three mechanisms for lymphocyte adhesion to cultured endothelial cells. *J. Cell Biol.* 107:321–331.
- Dustin, M. L., R. Rothlein, A. K. Bhan, C. A. Dinarello, and T. A. Springer. 1986. Induction by IL-1 and interferon, tissue distribution, biochemistry, and function of a natural adherence molecule (ICAM-1). *J. Immunol.* 137:245–254.
- Dustin, M. L., K. H. Singer, D. T. Tuck, and T. A. Springer. 1988. Adhesion of T lymphoblasts to epidermal keratinocytes is regulated by interferon gamma and is mediated by intercellular adhesion molecule-1 (ICAM-1). *J. Exp. Med.* 167:1323–1340.
- Girma, J.-P., D. Meyer, C. L. Verweij, H. Pannekoek, and J. J. Sixma. 1987. Structure-function relationship of human Von Willebrand factor. *Blood* 70:605–611.

16. Hardman, N. 1986. Structure and function of repetitive DNA in eukaryotes. *Biochem. J.* 234:1-11.
17. Haskard, D. O., D. Cavender, R. M. Fleck, R. Sontheimer, and M. Ziff. 1987. Human dermal microvascular endothelial cells behave like umbilical vein endothelial cells in T-cell adhesion studies. *J. Invest. Dermatol.* 88:340-344.
18. Henikoff, S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene.* 28:351-359.
19. Hopp, T. P., and K. R. Woods. 1988. Prediction of protein antigenic determinants from amino acid sequences. *Proc. Natl. Acad. Sci. USA.* 78:3824-3828.
20. Hunkapiller, M. W., E. Lujan, F. Ostrander, and L. E. Hood. 1983. Isolation of microgram quantities of proteins from polyacrylamide gels for amino acid sequence analysis. *Methods Enzymol.* 91:227-236.
21. Hynes, R. O. 1987. Integrins: a family of cell surface receptors. *Cell.* 48:549-554.
22. Kishimoto, T. K., N. Hollander, T. M. Roberts, D. C. Anderson, and T. A. Springer. 1987. Heterogenous mutations in the beta subunit common to the LFA-1, Mac-1, and p150.95 glycoproteins cause leukocyte adhesion deficiency. *Cell.* 50:193-202.
23. Kishimoto, T. K., K. O'Connor, A. Lee, T. M. Roberts, and T. A. Springer. 1987. Cloning of the beta subunit of the leukocyte adhesion proteins: homology to an extracellular matrix receptor defines a novel supergene family. *Cell.* 48:681-690.
24. Kozak, M. 1984. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucleic Acids Res.* 12:857-872.
25. Kürzinger, K., and T. A. Springer. 1982. Purification and structural characterization of LFA-1, a lymphocyte function-associated antigen, and Mac-1, a related macrophage differentiation antigen. *J. Biol. Chem.* 257:12412-12418.
26. Lathe, R. 1985. Synthetic oligonucleotide probes deduced from amino acid sequence data: theoretical and practical considerations. *J. Mol. Biol.* 183:1-12.
27. Law, S. K. A., J. Gagnon, J. E. K. Hildreth, C. E. Wells, A. C. Willis, and A. J. Wong. 1987. The primary structure of the beta subunit of the cell surface adhesion glycoproteins LFA-1, CR3 and p150.95 and its relationship to the fibronectin receptor. *EMBO (Eur. Mol. Biol. Organ.) J.* 6:915-919.
28. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. 545 pp.
29. Marlin, S. D., and T. A. Springer. 1987. Purified intercellular adhesion molecule-1 (ICAM-1) is a ligand for lymphocyte function-associated antigen 1 (LFA-1). *Cell.* 51:813-819.
30. Mentzer, S. J., R. Rothlein, T. A. Springer, and D. V. Faller. 1988. Intercellular adhesion molecule-1 (ICAM-1) is involved in the cytolytic T lymphocyte interaction with human synovial cells. *J. Cell. Physiol.* 137:173-178.
31. Messing, J. 1983. New M13 vectors for cloning. *Methods Enzymol.* 101:20-78.
32. Miller, L. J., and T. A. Springer. 1987. Biosynthesis and glycosylation of p150.95 and related leukocyte adhesion proteins. *J. Immunol.* 139:842-847.
33. Miller, L. J., M. Wiebe, and T. A. Springer. 1987. Purification and alpha subunit N-terminal sequences of human Mac-1 and p150.95 leukocyte adhesion proteins. *J. Immunol.* 138:2381-2383.
34. Mole, J. E., J. K. Anderson, E. A. Davison, and D. E. Woods. 1984. Complete primary structure for the zymogen of human complement factor B. *J. Biol. Chem.* 259:3407-3412.
35. Paulsson, M., and D. Heinegaard. 1984. Noncollagenous cartilage proteins. Current status of an emerging research field. *Collagen Relat. Res.* 4:219-229.
36. Pober, J. S., L. A. Lapierre, A. H. Stolpen, T. A. Brock, T. A. Springer, W. Fiers, M. P. Bevilacqua, D. L. Mendrick, and M. A. Gimbrone, Jr. 1987. Activation of cultured human endothelial cells by recombinant lymphotoxin: comparison with tumor necrosis factor and interleukin 1 species. *J. Immunol.* 138:3319-3324.
37. Poncz, M., R. Eisman, R. Heidenreich, S. M. Silver, G. Vileire, S. Surrey, E. Schwartz, and J. S. Bennett. 1987. Structure of the platelet membrane glycoprotein IIb; homology to the alpha subunits of the vitronectin and fibronectin membrane receptors. *J. Biol. Chem.* 262:8476-8482.
38. Pytela, R. 1988. Amino acid sequence of the murine Mac-1 alpha chain reveals homology with the integrin family and an additional domain related to von Willebrand factor. *EMBO (Eur. Mol. Biol. Organ.) J.* 7:1371-1378.
39. Rothlein, R., and T. A. Springer. 1986. The requirement for lymphocyte function-associated antigen 1 in homotypic leukocyte adhesion stimulated by phorbol ester. *J. Exp. Med.* 163:1132-1149.
40. Rothlein, R., M. L. Dustin, S. D. Marlin, and T. A. Springer. 1986. A human intercellular adhesion molecule (ICAM-1) distinct from LFA-1. *J. Immunol.* 137:1270-1274.
41. Ruoslahti, E., and M. D. Pierschbacher. 1987. New perspectives in cell adhesion. RGD and integrins. *Science (Wash. DC).* 238:491-497.
42. Sanchez-Madrid, F., J. Nagy, E. Robbins, P. Simon, and T. A. Springer. 1983. A human leukocyte differentiation antigen family with distinct α -subunits and a common β -subunit: the lymphocyte function-associated antigen (LFA-1), the C3bi complement receptor (OKM1/Mac-1), and the p150.95 molecule. *J. Exp. Med.* 158:1785-1803.
43. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 74:5463-5467.
44. Sastre, L., T. K. Kishimoto, C. Gee, T. Roberts, and T. A. Springer. 1986. The mouse leukocyte adhesion proteins Mac-1 and LFA-1: studies on mRNA translation and protein glycosylation with emphasis on Mac-1. *J. Immunol.* 137:1060-1065.
45. Shelton-Inloes, B. B., K. Titani, and J. E. Sadler. 1986. cDNA sequences for human von Willebrand factor reveal five types of repeated domains and five possible protein sequence polymorphisms. *Biochemistry.* 25:3164-3171.
46. Simmons, D., M. W. Makgoba, and B. Seed. 1988. ICAM, an adhesion ligand of LFA-1, is homologous to the neural cell adhesion molecule NCAM. *Nature (Lond.).* 331:624-627.
47. Springer, T. A., M. L. Dustin, T. K. Kishimoto, and S. D. Marlin. 1987. The lymphocyte function-associated LFA-1, CD2, and LFA-3 molecules: cell adhesion receptors of the immune system. *Annu. Rev. Immunol.* 5:223-252.
48. Springer, T. A., D. B. Teplow, and W. J. Dreyer. 1985. Sequence homology of the LFA-1 and Mac-1 leukocyte adhesion glycoproteins and unexpected relation to leukocyte interferon. *Nature (Lond.).* 314:540-542.
49. Staunton, D. E., S. D. Marlin, C. Stratowa, M. L. Dustin, and T. A. Springer. 1988. Primary structure of intercellular adhesion molecule 1 (ICAM-1) demonstrates interaction between members of the immunoglobulin and integrin supergene families. *Cell.* 52:925-933.
50. Suzuki, S., W. S. Argraves, H. Arai, L. R. Languino, M. D. Pierschbacher, and E. Ruoslahti. 1987. Amino acid sequence of the vitronectin receptor alpha subunit and comparative expression of adhesion receptor mRNAs. *J. Biol. Chem.* 262:14080-14085.
51. Suzuki, S., R. Pytela, H. Arai, W. S. Argraves, T. Krusius, M. D. Pierschbacher, and E. Ruoslahti. 1986. cDNA and amino acid sequences of the cell adhesion protein receptor recognizing vitronectin reveal a transmembrane domain and homologies with other adhesion protein receptors. *Proc. Natl. Acad. Sci. USA.* 83:8614-8618.
52. Szebenyi, D. M. E., S. K. Obendorf, and K. Moffat. 1981. Structure of vitamin D-dependent calcium-binding protein from bovine intestine. *Nature (Lond.).* 294:327-332.
53. te Velde, A. A., G. D. Keizer, and C. G. Figdor. 1987. Differential function of LFA-1 family molecules (CD11 and CD18) in adhesion of human monocytes to melanoma and endothelial cells. *Immunology.* 61:261-267.
54. von Heijne, G. 1984. How signal sequences maintain cleavage specificity. *J. Mol. Biol.* 173:243-251.
55. Vyas, N. K., M. N. Vyas, and F. A. Ouyoch. 1987. A novel calcium binding site in the galactose-binding protein of bacterial transport and chemotaxis. *Nature (Lond.).* 327:635-640.
56. Wilbur, W. J., and D. J. Lipman. 1983. Rapid similarity searches of nucleic acid and protein data banks. *Proc. Natl. Acad. Sci. USA.* 80:726-730.
57. Wong, W. W., L. B. Klickstein, J. A. Smith, J. H. Weis, and D. T. Fearon. 1985. Identification of a partial cDNA clone for the human receptor for complement fragment C3b/C4b. *Proc. Natl. Acad. Sci. USA.* 82:7711-7715.